The Hardness of Approximating the Boxicity, Cubicity and Threshold Dimension of a Graph

Abhijin Adiga[1], Diptendu Bhowmick[1], L. Sunil Chandran[1]

Department of Computer Science and Automation, Indian Institute of Science, Bangalore–560012, India.
emails: abhijin@csa.iisc.ernet.in, diptendubhowmick@gmail.com, sunil@csa.iisc.ernet.in

Abstract. A k-dimensional box is the Cartesian product $R_1 \times R_2 \times \cdots \times R_k$ where each R_i is a closed interval on the real line. The *boxicity* of a graph G, denoted as $\mathrm{box}(G)$, is the minimum integer k such that G can be represented as the intersection graph of a collection of k-dimensional boxes. A unit cube in k-dimensional space or a k-cube is defined as the Cartesian product $R_1 \times R_2 \times \cdots \times R_k$ where each R_i is a closed interval on the real line of the form $[a_i, a_i + 1]$. The *cubicity* of G, denoted as $\mathrm{cub}(G)$, is the minimum integer k such that G can be represented as the intersection graph of a collection of k-cubes. The *threshold dimension* of a graph $G(V, E)$ is the smallest integer k such that E can be covered by k threshold spanning subgraphs of G. In this paper we will show that there exists no polynomial-time algorithm to approximate the threshold dimension of a graph on n vertices with a factor of $O(n^{0.5-\epsilon})$ for any $\epsilon > 0$, unless $NP = ZPP$. From this result we will show that there exists no polynomial-time algorithm to approximate the boxicity and the cubicity of a graph on n vertices with factor $O(n^{0.5-\epsilon})$ for any $\epsilon > 0$, unless $NP = ZPP$. In fact all these hardness results hold even for a highly structured class of graphs namely the split graphs. We will also show that it is NP-complete to determine if a given split graph has boxicity at most 3.

Keywords: Boxicity, Cubicity, Threshold dimension, Partial order dimension, Split graph, NP-completeness, Approximation hardness

1 Introduction

Let $G(V, E)$ be a simple undirected finite graph with vertex set V and edge set E. A d-dimensional box is a Cartesian product $R_1 \times R_2 \times \cdots \times R_d$ where each R_i (for $1 \le i \le d$) is a closed interval of the form $[a_i, b_i]$ on the real line. A k-box representation of G is a mapping of the vertices of G to k-boxes such that two vertices in G are adjacent if and only if their corresponding k-boxes have a non-empty intersection. The *boxicity* of a graph denoted $\mathrm{box}(G)$, is the minimum integer k such that G can be represented as the intersection graph of k-dimensional boxes. A d-dimensional cube is a Cartesian product $R_1 \times R_2 \times \cdots \times R_d$ where each R_i (for $1 \le i \le d$) is a closed interval of the form $[a_i, a_i + 1]$

This work was supported by DST grant SR/S3/EECE/62/2006 and Infosys Fellowship.

on the real line. A k-cube representation of a graph G is a mapping of the vertices of G to k-cubes such that two vertices in G are adjacent if and only if their corresponding k-cubes have a non-empty intersection. The *cubicity* of G is the minimum integer k such that G has a k-cube representation.

The concept of boxicity was introduced by Roberts [10]. Cozzens [3] showed that computing the boxicity of a graph is NP-hard. This was later strengthened by Yannakakis [14] and finally by Kratochvíl [9] who showed that determining whether boxicity of a graph is at most two is NP-complete. In [14] Yannakakis showed that it is NP-complete to determine whether the cubicity of a given graph is at most 3.

1.1 Interval Graphs

A graph G is an *interval graph* if and only if G has an interval representation: i.e. each vertex of G can be associated with an interval on the real line such that two intervals intersect if and only if the corresponding vertices are adjacent. An interval graph G is said to be a *unit interval graph* if and only if there is some interval representation of G in which all the intervals are of the same length. Clearly, graphs with boxicity at most 1 are precisely the *interval graphs* and the graphs with cubicity at most 1 are precisely the *unit interval graphs*.

1.2 Split Graphs

A graph $G(V, E)$ is a split graph if its vertex set can be partitioned into a clique and an independent set. We will denote the clique by $\mathcal{C}(G)$ and independent set by $\mathcal{I}(G)$. Note that this partition need not be unique. But whenever we refer to $\mathcal{C}(G)$, the set $V \setminus \mathcal{C}(G)$ is an independent set and is denoted by $\mathcal{I}(G)$. Split graphs were first studied by Földes and Hammer in [6,2], and independently introduced by Tyshkevich and Chernyak [13]. For other characterizations and properties of split graphs one can refer to Golumbic [7].

Fact 1. *Complement of a split graph is a split graph.*

Definition 1. *A split interval graph is a graph which is both a split graph and an interval graph.*

1.3 Threshold graphs and the Threshold Dimension Problem

A graph is a threshold graph if there is a real number S and a weight function $w : V \longrightarrow \mathbb{R}$ such that for any two vertices $u, v \in V(G)$, (u, v) is an edge if and only if $w(u)+w(v) \geq S$. We will use the following property frequently in later sections.

Fact 2. *A graph $G(V, E)$ is a threshold graph if and only if it is a split graph and for every pair of vertices $u, v \in \mathcal{I}(G)$, either $N(u) \subseteq N(v)$ or $N(v) \subseteq N(u)$. Equivalently, a*

threshold graph can be defined as a split graph without an induced P_4 (i.e. a path on 4 vertices).

Note that threshold graphs are interval graphs.

Fact 3. *Complement of a threshold graph is a threshold graph.*

Definition 2. *Threshold dimension: A* threshold cover *of a graph G is a set of threshold graphs G_i, $i = 1, 2, \ldots, k$ on the same vertex set as G such that $E(G) = E(G_1) \cup E(G_2) \cup \cdots \cup E(G_k)$. The* threshold dimension *$t(G)$ is the least integer k such that a threshold cover of size k exists.*

Chvátal and Hammer [2] introduced threshold graphs and threshold dimension for their application in set-packing problems. In [14], Yannakakis showed that to determine if the threshold dimension of a graph is at most 3 is NP-complete even for the class of split graphs.

For a graph G let G_i, $1 \leq i \leq k$ be graphs on the same vertex set as G such that $E(G) = E(G_1) \cap E(G_2) \cap \cdots \cap E(G_k)$. Then we say that G is the *intersection* graph of G_i s for $1 \leq i \leq k$ and denote it as $G = \bigcap_{i=1}^{k} G_i$.

Fact 4. *From Fact 3 it is easy to see that threshold dimension of a graph G is the smallest integer k such that the complement graph $\overline{G}$ can be represented as the intersection of k threshold graphs. Also, if $G = \bigcap_{i=1}^{k} G_i$, then $t(\overline{G}) \leq \sum_{i=1}^{k} t(\overline{G_i})$.*

Lemma 1. *Let G be a split graph. Let G' be a threshold supergraph of G. Then we can construct another threshold graph H such that $G \subseteq H \subseteq G'$ and $\mathcal{I}(H) = \mathcal{I}(G)$.*

Proof. First we observe that $\mathcal{C}(G) \subseteq \mathcal{C}(G')$. The graph H is obtained as follows: $\mathcal{C}(H) = \mathcal{C}(G)$ and $\mathcal{I}(H) = \mathcal{I}(G)$. For each $u \in \mathcal{I}(H)$, $N(u, H) = N(u, G') \cap \mathcal{C}(G)$. By definition, $N(u, G) \subseteq N(u, H) \subseteq N(u, G')$. Therefore $G \subseteq H \subseteq G'$.

Now we will show that H is a threshold graph. Suppose there exist $u, v \in \mathcal{I}(H)$, such that neither $N(u, H) \subseteq N(v, H)$ nor $N(v, H) \subseteq N(u, H)$. There exist two vertices $u', v' \in \mathcal{C}(H)$ such that $u' \in N(u, H) \setminus N(v, H)$ and $v' \in N(v, H) \setminus N(u, H)$. This implies $u' \in N(u, G') \setminus N(v, G')$ and $v' \in N(v, G') \setminus N(u, G')$, which in turn implies that $u'uvv'$ forms an induced P_4 in G'. But, by Fact 2, this is a contradiction since G' is a threshold graph. $\square$

1.4 Posets

A *partially ordered set* (or *poset*) $P = (S, \leq_P)$ consists of a non-empty finite set S and a reflexive, antisymmetric and transitive binary relation $\leq_P$ on S. S is called the *ground set* of P. If $x \leq_P y$ or $y \leq_P x$ then x and y are said to be comparable. Otherwise we say

that they are incomparable and we denote this relation as $x||_P y$. We write $x <_P y$ when $x \leq_P y$ and $x \neq y$.

A *totally ordered set* is a poset in which every two elements are comparable. A *linear extension* L of a poset P is a totally ordered set $(S, \leq_L)$ which satisfies: $x \leq_P y \Longrightarrow x \leq_L y$. Let $L(u) = |\{v | v \leq_L u\}|$ denote the *index* of the element u in the totally ordered set L.

A *realizer* of a poset P is a set of linear extensions of P, say $\mathcal{L} : L_1, L_2, \ldots, L_k$ which satisfy the following condition: if $x||_P y$ then there exists two linear extensions $L_i, L_j \in \mathcal{L}$ such that $x <_{L_i} y$ and $y <_{L_j} x$. The *poset dimension* of P denoted by $\dim(P)$ is the minimum integer k such that there exists a realizer of P of cardinality k. It was introduced by Dushnik and Miller [5]. The poset dimension problem is to decide for a given poset and integer d whether the dimension of the poset is at most d. For a survey on dimension theory of posets see Trotter's monograph [11] or survey paper [12].

In [14] Yannakakis studied the complexity of the partial order dimension problem and its consequences on various graph parameters. He proved that it is NP-complete to determine whether the dimension of a partial order is at most 3. He then used some simple reductions to extend this result to the problems of determining the threshold dimension, boxicity and cubicity of graphs. Recently in [8] Hegde and Jain reduced the fractional chromatic number problem to the poset dimension problem to show that it is hard to even approximate the dimension of a partial order. To state more precisely,

Theorem 1. [8] *There exists no polynomial-time algorithm to approximate the poset dimension on an N-element set with a factor of $O(N^{0.5-\epsilon})$ for any $\epsilon > 0$, unless $NP = ZPP$.*

1.5 Our Results

In this paper we will show that

1. There exists no polynomial-time algorithm to approximate the threshold dimension of a graph on n vertices with a factor of $O(n^{0.5-\epsilon})$ for any $\epsilon > 0$, unless $NP = ZPP$.
2. There exists no polynomial-time algorithm to approximate the boxicity of a graph on n vertices with a factor of $O(n^{0.5-\epsilon})$ for any $\epsilon > 0$, unless $NP = ZPP$.
3. There exists no polynomial-time algorithm to approximate the cubicity of a graph on n vertices with a factor of $O(n^{0.5-\epsilon})$ for any $\epsilon > 0$, unless $NP = ZPP$.
4. If G is a split graph then it is NP-complete to determine whether $\text{box}(G) \leq 3$.

2 Preliminaries

Let G be a simple finite undirected graph on n vertices. The vertex set of G is denoted as $V(G)$ and the edge set of G is denoted as $E(G)$. For each vertex $v \in V(G)$ let $N(v, G)$

denote the set of vertices in $V(G)$ to which v is adjacent. Whenever there is no ambiguity regarding the graph under consideration, we will use the abbreviated notation $N(v)$. A graph H is said to be a subgraph of G if and only if $V(H) \subseteq V(G)$ and $E(H) \subseteq E(G)$. In this paper we will use the notation $H \subseteq G$ to denote H is a subgraph of G. Let $V' \subseteq V$. $G[V']$ denotes the induced subgraph of G on the vertex set V'. For a positive integer k, let $[k]$ denote the set $\{1, 2, \ldots, k\}$.

Suppose I is an interval graph. Let us consider an interval representation of I. Without loss of generality we can assume that the endpoints of each interval are integers. For any vertex u, let $l(u)$ and $r(u)$ denote the integers corresponding to the left endpoint and right endpoint respectively of the interval corresponding to u.

Property 1. Helly property of intervals: Suppose $A_1, A_2, \ldots, A_k$ is a finite set of intervals on the real line with pairwise non-empty intersection. Then there exists a common point of intersection for all the intervals i.e. $\bigcap_{i=1}^{k} A_i \neq \emptyset$.

Let $I_1, I_2, \cdots, I_k$ be k interval graphs (unit interval graphs) such that $G = \bigcap_{i=1}^{k} I_i$. Then $I_1, I_2, \cdots, I_k$ is called an interval (unit interval) representation of G. Boxicity can be stated in terms of intersection of interval graphs as follows:

Lemma 2. Roberts [10] *The boxicity of a graph G is the minimum positive integer b such that G can be represented as the intersection of b interval graphs. Moreover, if $G = \bigcap_{i=1}^{m} G_i$ for some graphs G_i then* $\mathrm{box}(G) \leq \sum_{i=1}^{m} \mathrm{box}(G_i)$.

Similarly cubicity can be stated in terms of intersection of unit interval graphs as follows:

Lemma 3. Roberts [10] *The cubicity of a graph G is the minimum positive integer b such that G is the intersection of b unit interval graphs. Moreover, if $G = \bigcap_{i=1}^{m} G_i$ for some graphs G_i then* $\mathrm{cub}(G) \leq \sum_{i=1}^{m} \mathrm{cub}(G_i)$.

The *boxicity problem* is defined to be the problem of computing the boxicity for a given graph G.

3 Characteristic Poset of a Split Graph

In this section, we will introduce the concept of the characteristic poset of a split graph and we will relate the threshold dimension and the boxicity of split graphs to the dimension of this poset.

Definition 3. *Let G be a split graph with $\mathcal{I}(G)$ and $\mathcal{C}(G)$ being the independent set and clique respectively. Let $\mathcal{X}(G) = \{N(u, G) | u \in \mathcal{I}(G)\}$. The characteristic poset of G is $P = (\mathcal{X}(G), \subseteq)$, i.e. the set of neighborhoods of the independent set vertices ordered by inclusion.*

Note that the characteristic poset is unique to a split graph and by Fact 2, we can infer that the characteristic poset is a totally ordered set if and only if the split graph is a threshold graph.

Theorem 2. *Let P be the characteristic poset of the split graph G. Then,* $\dim(P) \leq t(\overline{G})$.

Proof. Let $t(\overline{G}) = k$. Suppose $\mathcal{T} : T_1, T_2, \ldots, T_k$ is a set of threshold graphs such that $\bigcap_{i=1}^{k} T_i = G$. From each T_i, we will construct linear extension L_i of P such that $L_1, L_2, \ldots, L_k$ form a realizer of P.

From Lemma 1 we can assume that $\mathcal{I}(T_i) = \mathcal{I}(G)$ for $1 \leq i \leq k$. For each T_i let $\mathcal{X}(T_i) = \{N(u, T_i) | u \in \mathcal{I}(G)\}$. Consider the function $f_i : \mathcal{X}(G) \longrightarrow \mathcal{X}(T_i)$ where, for $X \in \mathcal{X}(G)$, $f_i(X)$ is the smallest subset in $\mathcal{X}(T_i)$ containing X. Note that f_i is well-defined: For each $X \in \mathcal{X}(G)$, there exists an $X' \in \mathcal{X}(T_i)$ such that $X \subseteq X'$ since T_i is a supergraph of G. Moreover, the smallest subset $f_i(X)$ is unique since $\mathcal{X}(T_i)$ is a totally ordered set with respect to set inclusion. We define L_i as follows: For any two distinct elements $X, Y \in \mathcal{X}(G)$,

1. If $f_i(X) \subset f_i(Y)$, then, $X <_{L_i} Y$.
2. If $f_i(X) = f_i(Y)$ and $X <_P Y$, then, $X <_{L_i} Y$.
3. If $f_i(X) = f_i(Y)$ and $X||_P Y$, then, we either make $X <_{L_i} Y$ or $Y <_{L_i} X$.

Since T_i is a threshold graph, we observe that

$$\begin{aligned} X \subseteq Y &\Longrightarrow f_i(X) \subseteq f_i(Y) \\ &\Longrightarrow X \leq_{L_i} Y \end{aligned}$$

Hence, L_is are linear extensions of P. Suppose $X||_P Y$, then there exist $u, v \in \mathcal{I}(G)$ such that $N(u, G) = X$ and $N(v, G) = Y$ and therefore there exist $u', v' \in \mathcal{C}(G)$ such that $u' \in N(u, G) \setminus N(v, G)$ and $v' \in N(v, G) \setminus N(u, G)$. Since $\bigcap_{i=1}^{k} T_i = G$, there exist two threshold graphs $T_j, T_l \in \mathcal{T}$ such that $u' \notin N(v, T_j)$ and $v' \notin N(u, T_l)$. This implies that $f_j(Y) \subset f_j(X)$ and $f_l(X) \subset f_l(Y)$. Therefore, $Y <_{L_j} X$ and $X <_{L_l} Y$. Hence, we have proved that L_is form a realizer of P. □

Lemma 4. *Let G be a split graph. Let G' be an interval supergraph of G. Then we can construct a split interval graph H such that $G \subseteq H \subseteq G'$ and $\mathcal{I}(H) = \mathcal{I}(G)$.*

Proof. Consider an interval representation of G' such that it satisfies the following two properties: (1) None of the intervals used is a single point interval. (2) No two intervals share a common end point. It is easy to see that such an interval representation can be constructed from any given interval representation in polynomial time. Now let $x \in \mathcal{I}(G)$. Clearly $\{x\} \cup N(x, G)$ induces a clique in G and therefore in G'. Let $f'(v)$ denote the

interval assigned to the vertex v in the interval representation chosen for G'. By Helly property of the intervals, $\bigcap_{v \in \{x\} \cup N(x,G)} f'(v) \neq \emptyset$. From properties (1) and (2) we can easily infer that $\bigcap_{v \in \{x\} \cup N(x,G)} f'(v)$ is not a single point interval. Now we define the interval graph H on the vertex set $V(G)$, by assigning the interval $f(v)$ to each vertex $v \in V(G)$, defined as follows

$$f(v) = \begin{cases} f'(v) \ \forall v \in \mathcal{C}(G), \\ P(v) \ \forall v \in \mathcal{I}(G), \end{cases}$$

where $P(v)$ is a point in $\bigcap_{x \in \{v\} \cup N(v,G)} f'(x)$. Note that since $\bigcap_{x \in \{v\} \cup N(v,G)} f'(x)$ is not a single point we can assume that $P(v) \neq P(u)$ for all distinct $u, v \in \mathcal{I}(G)$. Also note that for each $v \in \mathcal{I}(G)$, $N(v,G) \subseteq N(v,H)$ by the construction. Since we have only changed the intervals corresponding to the vertices in $\mathcal{I}(G)$, we infer that $G \subseteq H$. On the other hand $f'(v) \supseteq f(v)$ for all $v \in V(G)$ and therefore $H \subseteq G'$, as required. Moreover it is easy to see that $\mathcal{I}(G)$ induces an independent set in H. Hence, H is a split graph with the same partition as G. Therefore, H is a split interval graph. □

Lemma 5. *If G is a split interval graph, then $t(\overline{G}) \leq 2$.*

Proof. Let us consider an interval representation of G. We will construct two threshold graphs G_1 and G_2 as follows. Let $l = \min_{u \in V(G)} l(u)$ and $r = \max_{u \in V(G)} r(u)$ be the leftmost and the rightmost points respectively, in the interval representation of G. Now, to define G_1, we change the intervals corresponding to $u \in \mathcal{C}(G)$ by redefining their left end points: $l(u) = l$, $\forall u \in \mathcal{C}(G)$. We do not disturb the intervals corresponding to the vertices in $\mathcal{I}(G)$. Now we claim that G_1 is a threshold graph: Clearly $\mathcal{I}(G)$ induces an independent set in G_1 also. Therefore let $\mathcal{I}(G_1) = \mathcal{I}(G)$. Let $u, v \in \mathcal{I}(G_1)$. It is easy to see that $N(u, G_1) \supseteq N(v, G_1)$ if $l(u) \leq l(v)$ and therefore, for every $u, v \in \mathcal{I}(G_1)$, we have either $N(u, G_1) \subseteq N(v, G_1)$ or $N(v, G_1) \subseteq N(u, G_1)$.

Similarly, let G_2 be obtained by letting $r(u) = r$, $\forall u \in \mathcal{C}(G)$, while keeping other end points unchanged. Again by construction, G_2 is a threshold graph. It is easy to see that $G_1 \cap G_2 = G$: By construction, $G_1 \supseteq G$ and $G_2 \supseteq G$ and if $(u, v) \notin E(G)$, it is clear that in G_1 or in G_2, the intervals corresponding to u and v are disjoint. □

Lemma 6. *If G is a split graph, then $t(\overline{G}) \leq 2\mathrm{box}(G)$.*

Proof. Let $\mathrm{box}(G) = k$ and $G_1, G_2, \ldots, G_k$ be interval graphs on the same vertex set as G such that $\bigcap_{i=1}^{k} G_i = G$. By Lemma 4, we can assume that all the G_is are split interval graphs. By Lemma 5, corresponding to each G_i, we can construct two threshold graphs T_{2i-1} and T_{2i} such that $G_i = T_{2i-1} \cap T_{2i}$. Therefore, we have $2k$ threshold graphs whose intersection gives G. Hence, proved. □

Combining the above Lemma and Theorem 2, we have:

Theorem 3. *Let $P = (S, \leq_P)$ be a characteristic poset of the split graph G. Then $\dim(P) \leq 2\text{box}(G)$.*

Remark 1. We observe that the constructions in Theorem 2 and Lemmas 4, 5 and 6 can be achieved in polynomial time.

4 Hardness of Approximation

Given poset P, we will construct a split graph G_P such that P is isomorphic to the characteristic poset of G_P. Consider a poset $P = (S, \leq_P)$ where $|S| = n$. Let $g : [n] \longrightarrow S$ be a bijective map. For convenience, we will assume that S and $[n]$ are disjoint sets. We define a split graph G_P as follows: $V(G_P) = S \cup [n]$. $\mathcal{C}(G_P) = [n]$ and $\mathcal{I}(G_P) = S$. For any $u \in S$ and $v \in [n]$, $(u, v) \in E(G_P) \iff g(v) \leq_P u$. Thus $g(N(u, G_P)) = \{x \in S | x \leq_P u\}$. It is easy to see that P is isomorphic to the characteristic poset of G_P.

Theorem 4. $\dim(P) \geq t(\overline{G_P})$.

Proof. Let $\dim(P) = k$. Suppose $L_1, L_2, \ldots, L_k$ form a realizer of P. We will construct threshold graphs G_i corresponding to each L_i for $1 \leq i \leq k$ such that $\bigcap_{i=1}^{k} G_i = G_P$. The G_is are defined as follows: $V(G_i) = S \cup [n]$ with $\mathcal{C}(G_i) = [n]$ and $\mathcal{I}(G_i) = S$. For any $u \in S$ and $v \in [n]$, $(u, v) \in E(G_i) \iff g(v) \leq_{L_i} u$. G_i is a threshold graph because L_i (a totally ordered set) is the characteristic poset of G_i.

Now, we will show that if $(u, v) \in E(G_P)$ then $(u, v) \in E(G_i)$ $\forall i \in [k]$. Since $\mathcal{C}(G_i) = \mathcal{C}(G_P)$, any $u, v \in \mathcal{C}(G_i)$ are adjacent in G_i. Suppose $u \in \mathcal{I}(G_P)$ and $v \in \mathcal{C}(G_P)$,

$$\begin{aligned}
(u, v) \in E(G_P) &\Longrightarrow g(v) \leq_P u \\
&\Longrightarrow g(v) \leq_{L_i} u, \forall i \in [k] \\
&\Longrightarrow (u, v) \in E(G_i), \forall i \in [k]
\end{aligned}$$

Hence, each G_i is a supergraph of G_P. Next we will show that if $(u, v) \notin E(G_P)$ then there exists G_j such that $(u, v) \notin E(G_j)$. If $(u, v) \notin E(G_P)$ then either $u <_P g(v)$ or $u ||_P g(v)$. In either case, there exists an L_j such that $u <_{L_j} g(v)$. By definition of G_j, $(u, v) \notin E(G_j)$. Hence, proved. □

Combining Theorems 2 and 4, we have the following result.

Corollary 1. $\dim(P) = t(\overline{G_P})$.

Cozzens and Halsey [4] proved that the boxicity of any graph $G(V, E)$ is not more than the threshold dimension of it's complement $\overline{G}$, i.e. $\text{box}(G) \leq t(\overline{G})$. Hence,

Corollary 2. $\dim(P) \geq \text{box}(G_P)$.

Remark 2. We note that the construction in Theorem 4 can be achieved in polynomial time.

Theorem 5. *There exists no polynomial-time algorithm to approximate the threshold dimension of a split graph on n vertices with a factor of $O(n^{0.5-\epsilon})$ for any $\epsilon > 0$ unless $NP = ZPP$.*

Proof. Suppose there exists an algorithm to compute the boxicity of a split graph on n vertices with approximation factor $O(n^{0.5-\epsilon})$. As we have seen for any poset P on N elements we can construct a split graph G_P on $n = 2N$ vertices such that $t(\overline{G_P}) = \dim(P)$ by Corollary 1. This immediately implies that $\dim(P)$ can be approximated within factor $O(n^{0.5-\epsilon})$. But, from Theorem 1 we know that there exists no polynomial-time algorithm to approximate the poset dimension problem with a factor $O(n^{0.5-\epsilon})$ for any $\epsilon > 0$ unless $NP = ZPP$, a contradiction. □

Theorem 6. *There exists no polynomial-time algorithm to approximate the boxicity of a split graph on n vertices with a factor of $O(n^{0.5-\epsilon})$ for any $\epsilon > 0$ unless $NP = ZPP$.*

Proof. The proof is similar to that of Theorem 5. From Theorem 3 and Corollary 2, we have $\mathrm{box}(G_P) \leq \dim(P) \leq 2\mathrm{box}(G_P)$. The rest follows from Theorem 1. □

Corollary 3. *There exists no polynomial-time algorithm to approximate the cubicity of a split graph on n vertices with a factor of $O(n^{0.5-\epsilon})$ for any $\epsilon > 0$ unless $NP = ZPP$.*

Proof. In [1] it is shown that for any graph G on n vertices, $\mathrm{cub}(G) \leq \mathrm{box}(G) \lceil \log_2 n \rceil$. Since any representation of G as the intersection of cubes also serves as an intersection of boxes, it follows that $\mathrm{cub}(G) \geq \mathrm{box}(G)$. Hence, given a poset P and the corresponding split graph G_P as constructed in Section 4, we have $\mathrm{cub}(G_P)/\lceil \log_2 n \rceil \leq \dim(P) \leq 2\mathrm{cub}(G_P)$. The rest follows as in Theorem 5. □

5 NP-Completeness of Boxicity of Split Graph

The following theorem was proved by Yannakakis in [14].

Theorem 7. [14] *It is NP-complete to determine if a given split graph has threshold dimension at most 3.*

We will reduce the threshold dimension problem of split graphs to the problem of computing boxicity of a split graph. Let H be any split graph. Let $|V(H)| = n$. We will construct another split graph G' in polynomial time such that $\mathrm{box}(G') = t(H)$. A split graph G is said to be a complete split graph if for all $u \in \mathcal{I}(G)$ and $v \in \mathcal{C}(G)$, $(u, v) \in E(G)$. Note that a complete split graph is also a threshold graph. If H is a complete split graph then we take $G' = H$ since $\mathrm{box}(H) = t(H) = 1$. So for the rest of the proof we will

assume that H is not a complete split graph. Let $G = \overline{H}$ and G_1, G_2 be copies of G. Let $V(G_1) = \mathcal{C}(G_1) \cup \mathcal{I}(G_1)$ and $V(G_2) = \mathcal{C}(G_2) \cup \mathcal{I}(G_2)$. $V(G') = V(G_1) \cup V(G_2)$ and $E(G') = E(G_1) \cup E(G_2) \cup \{(u,v) | u \in \mathcal{C}(G_1), v \in \mathcal{C}(G_2)\} \cup \{(u,v) | u \in \mathcal{C}(G_1), v \in \mathcal{I}(G_2)\} \cup \{(u,v) | u \in \mathcal{C}(G_2), v \in \mathcal{I}(G_1)\}$. Clearly, G' is a split graph with $\mathcal{C}(G') = \mathcal{C}(G_1) \cup \mathcal{C}(G_2)$.

5.1 $\mathrm{box}(\boldsymbol{G'}) \leq \boldsymbol{t(H)}$

Let $t(H) = k$ and $T_1, T_2, \ldots, T_k$ be a set of threshold graphs such that $\bigcap_{i=1}^{k} T_i = G$. Due to Lemma 1, we can assume that $\mathcal{I}(T_i) = \mathcal{I}(G)$. Now we construct interval graphs H_i corresponding to each T_i as follows: Let T_i^1 and T_i^2 be two copies of T_i. We assume that $V(G_1) = V(T_i^1)$ and $V(G_2) = V(T_i^2)$. Let $V(H_i) = V(G_1) \cup V(G_2)$. Let $g_i : \mathcal{I}(T_i^j) \longrightarrow [n]$, $j = 1, 2$, be a function which assigns to each vertex in the independent set of T_i^j a distinct number satisfying: $u, v \in \mathcal{I}(T_i^j)$, $N(u, T_i^j) \subset N(v, T_i^j) \implies g_i(u) > g_i(v)$. We define another function $h_i : \mathcal{C}(T_i^j) \longrightarrow [n]$, $j = 1, 2$, as: $\forall u \in \mathcal{C}(T_i^j)$

$$h_i(u) = \begin{cases} 0, & \text{if } N(u, T_i^j) \cap \mathcal{I}(T_i^j) = \varnothing, \\ \max\limits_{v \in N(u, T_i^j) \cap \mathcal{I}(T_i^j)} g_i(v), & \text{otherwise.} \end{cases}$$

Each $u \in \mathcal{I}(T_i^1)$ is associated with the single point interval $[g_i(u), g_i(u)]$ and $u \in \mathcal{C}(T_i^1)$ with interval $[-n, h_i(u)]$. Each $u \in \mathcal{I}(T_i^2)$ is associated with the single point interval $[-g_i(u), -g_i(u)]$ and $u \in \mathcal{C}(T_i^2)$ with interval $[-h_i(u), n]$. Now H_i is defined to be the intersection graph of this family of intervals which corresponds to $V(G_1) \cup V(G_2)$.

Remark 3. $\mathcal{C}(T_i^j) = \mathcal{C}(G_j)$ and $\mathcal{I}(T_i^j) = \mathcal{I}(G_j)$ for $1 \leq i \leq k$ and $j = 1, 2$.

Lemma 7. *H_i is a split graph with $\mathcal{C}(H_i) = \mathcal{C}(G')$ and $\mathcal{I}(H_i) = \mathcal{I}(G')$ for $1 \leq i \leq k$.*

Proof. In view of the construction of H_i clearly, 0 is a common point for intervals corresponding to all vertices $u \in \mathcal{C}(T_i^1) \cup \mathcal{C}(T_i^2)$. Also, by definition of g_i, it follows that intervals corresponding to all vertices $u \in \mathcal{I}(T_i^1) \cup \mathcal{I}(T_i^2)$ are mutually disjoint. Hence, $\mathcal{C}(H_i) = \mathcal{C}(G')$ and $\mathcal{I}(H_i) = \mathcal{I}(G')$. Therefore, H_i is a split graph. □

Lemma 8. *$H_i[V(G_1)] = T_i^1$ and $H_i[V(G_2)] = T_i^2$ for $1 \leq i \leq k$.*

Proof. Clearly $H_i[V(G_1)]$ is a split graph with $\mathcal{I}(H_i[V(G_1)]) = \mathcal{I}(T_i^1)$ and $\mathcal{C}(H_i[V(G_1)]) = \mathcal{C}(T_i^1)$. By construction it is easy to see that $E(H_i[V(G_1)]) \supseteq E(T_i^1)$. Let $x \in \mathcal{I}(T_i^1)$ and $y \in \mathcal{C}(T_i^1)$ such that $(y, x) \notin E(T_i^1)$. Let $z \in \mathcal{I}(T_i^1)$ be such that $(y, z) \in E(T_i^1)$. According to Fact 2 we have either $N(x, T_i^1) \subseteq N(z, T_i^1)$ or $N(x, T_i^1) \supseteq N(z, T_i^1)$. But since $y \notin N(x, T_i^1)$ and $y \in N(z, T_i^1)$ we can infer that $N(x, T_i^1) \subset N(z, T_i^1)$. It follows that $g_i(x) > g_i(z)$. Clearly $h_i(y) \leq g_i(z) < g_i(x)$. Therefore $(x, y) \notin E(H_i[V(G_1)])$ and therefore $H_i[V(G_1)] = T_i^1$. A similar proof shows that $H_i[V(G_2)] = T_i^2$. □

Lemma 9. *$\mathrm{box}(G') \leq t(H)$.*

Proof. According to Lemma 7, $\mathcal{C}(H_i) = \mathcal{C}(G')$ and $\mathcal{I}(H_i) = \mathcal{I}(G')$ for $1 \leq i \leq k$. Let $u \in \mathcal{C}(G')$ and $v \in \mathcal{I}(G')$. We consider the following cases:

1. $u \in \mathcal{C}(G_1)$ and $v \in \mathcal{I}(G_2)$: Then $(u, v) \in E(G')$ by construction of G'. According to Remark 3 and by construction of H_i, the interval corresponding to $u \in \mathcal{C}(T_i^1)$ contains $[-n, 0]$ and $v \in \mathcal{I}(T_i^2)$ corresponds to a single point interval on the negative x-axis. It follows that $(u, v) \in E(H_i)$ for $1 \leq i \leq k$.
2. $u \in \mathcal{C}(G_2)$ and $v \in \mathcal{I}(G_1)$: Similar to case 1.
3. $u \in \mathcal{C}(G_1)$ and $v \in \mathcal{I}(G_1)$: Note that $G'[V(G_1)] = G_1$ and by Lemma 8, $H_i[V(G_1)] = T_i^1$ for $1 \leq i \leq k$. Since $\bigcap_{i=1}^{k} T_i^1 = G_1$ we have $\bigcap_{i=1}^{k} H_i[V(G_1)] = \bigcap_{i=1}^{k} T_i^1 = G_1 = G'[V(G_1)]$.
4. $u \in \mathcal{C}(G_2)$ and $v \in \mathcal{I}(G_2)$: Similar to case 3. We can show that $\bigcap_{i=1}^{k} H_i[V(G_2)] = \bigcap_{i=1}^{k} T_i^2 = G_2 = G'[V(G_2)]$.

From the above points we can infer that if $(u, v) \in E(G')$ then $(u, v) \in E(H_i)$ for $1 \leq i \leq k$ and if $(u, v) \notin E(G')$ then $(u, v) \notin E(H_l)$ for some $l \in [k]$. Therefore $\bigcap_{i=1}^{k} H_i = G'$ and hence $\mathrm{box}(G') \leq k = t(H)$. $\square$

5.2 $\mathrm{box}(G') \geq t(H)$

Let $\mathrm{box}(G') = l$ and $I_1, I_2, \ldots, I_l$ be interval graphs such that $\bigcap_{i=1}^{l} I_i = G'$. From Lemma 4 we can assume that each I_i is a split graph with $\mathcal{I}(I_i) = \mathcal{I}(G')$. Moreover,

Remark 4. $I_i[V(G_1)]$ and $I_i[V(G_2)]$ are split graphs with $\mathcal{I}(I_i[V(G_1)]) = \mathcal{I}(G_1)$ and $\mathcal{I}(I_i[V(G_2)]) = \mathcal{I}(G_2)$ respectively for $1 \leq i \leq l$.

We shall use the notation T_C to denote a complete split graph.

Lemma 10. *With respect to an interval representation of I_i, let u_l and u_r be the vertices corresponding to the leftmost and rightmost intervals respectively, among the vertices in $\mathcal{I}(I_i)$.*

1. *If $u_l \in \mathcal{I}(G_1)$ and $u_r \in \mathcal{I}(G_2)$ then $t(\overline{I_i[V(G_1)]}) = 1$ and $t(\overline{I_i[V(G_2)]}) = 1$.*
2. *If $u_l \in \mathcal{I}(G_2)$ and $u_r \in \mathcal{I}(G_1)$ then $t(\overline{I_i[V(G_1)]}) = 1$ and $t(\overline{I_i[V(G_2)]}) = 1$.*
3. *If $u_l, u_r \in \mathcal{I}(G_1)$ then $t(\overline{I_i[V(G_1)]}) \leq 2$ and $I_i[V(G_2)] = T_C$.*
4. *If $u_l, u_r \in \mathcal{I}(G_2)$ then $I_i[V(G_1)] = T_C$ and $t(\overline{I_i[V(G_2)]}) \leq 2$.*

Proof(1): First we will prove that $I_i[V(G_1)]$ is a threshold graph, which, by Fact 3, implies $t(\overline{I_i[V(G_1)]}) = 1$. By assumption $r(u) < r(u_r)$ for all $u \in \mathcal{I}(I_i)$, $u \neq u_r$. Since $\mathcal{I}(G_1) \cup \mathcal{I}(G_2)$ induces an independent set in I_i we have $r(u) < l(u_r)$ for all $u \in \mathcal{I}(G_1)$ because otherwise $l(u_r) \leq r(u) < r(u_r)$ and hence intervals corresponding to u and u_r intersect in the interval representation of I_i. For any $v \in \mathcal{C}(G_1)$, $r(v) \geq l(u_r)$ since by construction of G', $(v, u_r) \in E(G')$ and $G' \subseteq I_i$. Combining these two observations, we get $r(u) < l(u_r) \leq r(v)$ and thus $r(u) < r(v)$ for all $u \in \mathcal{I}(G_1), v \in \mathcal{C}(G_1)$. Suppose

$u_1, u_2 \in \mathcal{I}(G_1)$ such that $r(u_1) \leq r(u_2)$. Now for all $v \in \mathcal{C}(G_1)$, $r(u_1) \leq r(u_2) < r(v)$. If $(u_1, v) \in E(I_i[V(G_1)])$ then $l(v) \leq r(u_1) \leq r(u_2) < r(v)$. Hence $(u_2, v) \in E(I_i[V(G_1)])$ also. From this and Remark 4, it is clear that Fact 2 holds for $I_i[V(G_1)]$. Therefore $I_i[V(G_1)]$ is a threshold graph. Similarly, we can show that $t(\overline{I_i[V(G_2)]}) = 1$.
Proof(2): Similar to Proof of (1).
Proof(3): Since $\mathcal{I}(G_1) \cup \mathcal{I}(G_2)$ induces an independent set in I_i, we have for all $u \in \mathcal{I}(G_2)$, $l(u) > r(u_l)$ and $r(u) < l(u_r)$. Since by construction of G' for all $v \in \mathcal{C}(G_2)$, $(v, u_l) \in E(G')$, $(v, u_r) \in E(G')$ and $G' \subseteq I_i$, we have $l(v) \leq r(u_l)$ and $r(v) \geq l(u_r)$. This implies $l(v) < l(u) \leq r(u) < r(v)$ for all $u \in \mathcal{I}(G_2), v \in \mathcal{C}(G_2)$. Hence all vertices in $\mathcal{I}(G_2)$ are adjacent to all vertices in $\mathcal{C}(G_2)$. Now $I_i[V(G_2)]$ is a complete split graph and hence $I_i[V(G_2)] = T_C$. On the other hand by Remark 4, $I_i[V(G_1)]$ is a split interval graph. Hence from Lemma 5, $t(\overline{I_i[V(G_1)]}) \leq 2$.
Proof(4): Similar to Proof of (3).

Remark 5. Suppose G is a split graph with $t(\overline{G}) = k$. Let $\mathcal{T} : T_1, T_2, \ldots, T_k$ be a set of threshold graphs such that $\bigcap_{i=1}^{k} T_i = G$. It is easy to see that there does not exist a pair of graphs $T_i, T_j \in \mathcal{T}$ such that $T_i \subseteq T_j$. Suppose this was not the case, then, $G = \bigcap_{l=1, l \neq j}^{k} T_l$, i.e. we could discard T_j, thus contradicting the minimality of k.

Lemma 11. $\operatorname{box}(G') \geq t(H)$.

Proof. Based on Lemma 10, we can infer that $I_i[V(G_1)]$ belongs to exactly one of the following 3 cases: 1) $t(\overline{I_i[V(G_1)]}) = 1$ and $I_i[V(G_1)] \neq T_C$. 2) $t(\overline{I_i[V(G_1)]}) \leq 2$. 3) $I_i[V(G_1)] = T_C$. Let l_1, l_2, l_3 be such that l_j denotes the number of times $I_i[V(G_1)]$ belongs to case j for $1 \leq i \leq l$ and $1 \leq j \leq 3$. Clearly $l_1 + l_2 + l_3 = l$. Recall that H is not a complete split graph. Therefore there exists some $i \in [l]$ such that $I_i \neq T_C$. Note that $G_1 = \bigcap_{i=1}^{l} I_i[V(G_1)]$ and therefore $t(\overline{G_1}) \leq \sum_{i=1}^{l} t(\overline{I_i[V(G_1)]}) \leq l_1 + 2l_2 + l_3 t(\overline{T_C})$. Since any threshold graph T which is a supergraph of $\overline{H}$ is a subgraph of T_C, by Remark 5, T_C can be discarded and therefore, we can ignore the term $l_3 t(\overline{T_C})$ in the above expression. Hence we get $t(\overline{G_1}) \leq l_1 + 2l_2$.
We can get 3 similar cases for $I_i[V(G_2)]$. Let l'_j denotes the number of times $I_i[V(G_2)]$ belongs to case j for $1 \leq i \leq l$ and $1 \leq j \leq 3$. Clearly $l'_1 + l'_2 + l'_3 = l$. From Lemma 10, it is easy to see that $l'_1 = l_1$, $l'_2 = l_3$ and $l'_3 = l_2$. Therefore $t(\overline{G_2}) \leq \sum_{i=1}^{l} t(\overline{I_i[V(G_2)]})$ $\leq l_1 + 2l_3$. Hence realizing that $\overline{G_1}$ and $\overline{G_2}$ are isomorphic to H,

$$2t(H) = t(\overline{G_1}) + t(\overline{G_2}) \leq 2(l_1 + l_2 + l_3) = 2l.$$

Hence, we get $t(H) \leq l = \operatorname{box}(G')$. □

Theorem 8. *It is NP-complete to determine if a given split graph has boxicity at most 3.*

Proof. We reduce the problem of determining the threshold dimension of a split graph to this problem. Given a split graph H we can construct another split graph G' in polynomial time such that $\text{box}(G') = t(H)$ by Lemma 9 and Lemma 11. The rest follows from Theorem 7. $\square$

References

1. L. S. Chandran, K. A. Mathew, An upper bound for cubicity in terms of boxicity, Disc. Math. 309 (8) (2009) 2571–2574.
2. V. Chvátal, P. L. Hammer, Aggregation of inequalities in integer programming, in: Ann. Discrete Math, 1977.
3. M. B. Cozzens, Higher and multi-dimensional analogues of interval graphs, Ph.D. thesis, Department of Mathematics, Rutgers University, New Brunswick, NJ (1981).
4. M. B. Cozzens, M. D. Halsey, The relationship between the threshold dimension of split graphs and various dimensional parameters, Disc. Appl. Math. 30 (1991) 125–135.
5. B. Dushnik, E. W. Miller, Partially ordered sets, Amer. J. Math 6 (3) (1941) 600–610.
6. S. Foldes, P. L. Hammer, Split graphs, in: Proceedings of the 8th South-Eastern Conference on Combinatorics, Graph Theory and Computing, 1977.
7. M. C. Golumbic, Algorithmic Graph Theory and Perfect Graphs, Academic Press, New York, 1980.
8. R. Hegde, K. Jain, The hardness of approximating poset dimension, Electronic Notes on Discrete Mathematics 29 (2007) 435–443.
9. J. Kratochvíl, A special planar satisfiability problem and a consequence of its NP-completeness, Disc. Appl. Math. 52 (1994) 233–252.
10. F. S. Roberts, Recent Progresses in Combinatorics, chap. On the boxicity and cubicity of a graph, Academic Press, New York, 1969, pp. 301–310.
11. W. T. Trotter, Combinatorics and partially ordered sets: Dimension Theory, The Johns Hopkins University Press, Baltimore, Maryland, 1992.
12. W. T. Trotter, Graphs and partially ordered sets: recent results and new directions, in: Surveys in graph theory (San Fransisco, CA, 1995), Congr. Numer. 116, 1996.
13. R. I. Tyshkevich, A. A. Chernyak, Canonical partition of a graph defined by the degrees of its vertices, in: (In Russian) Isv. Akad. Nauk BSSR, Ser. Fiz.-Mat. Nauk, 1979.
14. M. Yannakakis, The complexity of the partial order dimension problem, SIAM J. Alg. Disc. Math. 3 (3) (1982) 351–358.